Computers in Human Behavior xxx (2018) 1–10

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Full length article

Bridging learning analytics and Cognitive Computing for Big Data classification in micro-learning video collections

Danilo Dessì, Gianni Fenu, Mirko Marras*, Diego Reforgiato Recupero

Department of Mathematics and Computer Science, University of Cagliari, Via Ospedale 72, 09124 Cagliari, Italy

A R T I C L E I N F O

Article history:
Received 10 February 2017
Received in revised form
19 July 2017
Accepted 1 March 2018
Available online xxx

Keywords:
Cognitive Computing
Big Data technologies
Micro-learning video
Multi-class classification
Learning Analytics
Video classification

A B S T R A C T

Moving towards the next generation of personalized learning environments requires intelligent approaches powered by analytics for advanced learning contexts with enriched digital content. Micro-Learning through Massive Open Online Courses is riding the wave of popularity as a novel paradigm for delivering short educational videos in small pre-organized chunks over time, so that learners can get knowledge in a manageable way. However, with the ever-increasing number of videos, it has become challenging to arrange and search them according to specific categories. In this paper, we get around the problem by bridging Learning Analytics and Cognitive Computing to analyze the content of large video collections, going over traditional term-based methods. We propose an efficient and effective approach to automatically classify a collection of educational videos on pre-existing categories which uses (i) a Speech-to-Text tool to get video transcripts, (ii) Natural Language Processing and Cognitive Computing methods to extract semantic concepts and keywords from video transcripts for their representation, and (iii) Apache Spark as Big Data technology for scalability. Several classifiers are trained on the feature vectors extracted by Cognitive Computing tools. Then, we compared our approach with other combinations of state-of-the-art feature types and classifiers over a large-scale dataset we collected from Coursera. Considering the experimental results, we expect our approach can facilitate the development of Learning Analytics tools powered by Cognitive Computing to support content managers on micro-learning video management while improving how learners search videos.

1. Introduction

The ever-increasing availability of digital data brings unprecedented possibilities to analyze different educational facets. Learners, teachers, designers and managers have been mobilized to investigate how data can be used to support learning and teaching. This intense interest has given rise to tools and techniques in the research field of Learning Analytics (LA) (Baker & Inventado, 2014). Notable examples include the identification of learners at risk of failing (Xing, Chen, Stein, & Marcinkowski, 2016) and the analysis of data flows coming from the interactions among communities (Barak, Watted, & Haick, 2016). The field is in its dynamic youth, and there are numerous opportunities to unlock the potential of LA, especially with the rapid development of emerging technologies such as Cognitive Computing and Big Data.

In the large set of educational data, micro-learning videos embedded into Massive Open Online Courses (MOOCs) represent one of the most powerful medium to deliver knowledge to learners in small chunks over time. This combination promises to have a solid future as proved by the recent experiences in Kloos, Jermann, Pérez-Sanagustín, Seaton, & White, 2017, as an example. Even more, over 58 million people were registered for a MOOC and about 6850 MOOCs were provided by over 700 institutions up to 2016.[1] In this context, the most popular providers are currently Coursera[2] and EdX.[3] In contrast with traditional educational models where learners interrupt their activities for hours, micro-learning processes from 5 to 15 min have a positive impact in knowledge acquisition (Coakley, Garvey, & ONeill, 2017). They provide

* Corresponding author.
E-mail addresses: danilo_dessi@unica.it (D. Dessì), fenu@unica.it (G. Fenu), mirko.marras@unica.it (M. Marras), diego.reforgiato@unica.it (D. Reforgiato Recupero).

1 https://www.class-central.com/report/mooc-stats-2016/.
2 https://www.coursera.org/.
3 https://www.edx.org/.

https://doi.org/10.1016/j.chb.2018.03.004
0747-5632/

adaption to the constraints of human brain with respect to attention span and high flexibility (Leach & Hadi, 2017; Sun et al., 2015).

However, with the ever-increasing number of micro-learning videos, managing large video collections is turning into a challenge for content managers. Human annotation and organization is time-consuming and non cost-effective. On the other hand, searching videos given specific categories of interest is becoming hard for learners. This has raised the need of LA tools powered by intelligent methods for effective and efficient video classification. Recent studies (Basu, Yu, & Zimmermann, 2016; Valiente, Sicilia, Garcia-Barriocanal, & Rajabi, 2015) tend to model the problem as a text classification task which automatically assigns one category from a set of predefined ones to a video based on the transcript (i.e. a written version of the content presented by the speaker). These works focused on traditional long face-to-face video lessons which provide a lot of information in comparison to micro-learning videos. Furthermore, these approaches usually map video transcripts using Term-Frequency-Inverse Document Frequencies (TF-IDF), one of the most popular representations in literature. Even if it is simple and widely-used, several limitations have emerged. Text documents are modeled as a set of term frequencies, regardless the position in the document or semantic links with other words. It follows that the knowledge derived by the information behind the text is lost.

Emerging Semantic Web resources and techniques have been recently combined with Data Mining and Knowledge Discovery methods in order to manage huge amount of data to perform analysis and obtain useful insights out of the data (Ristoski & Paulheim, 2016). Cutting-edge cognitive computing systems such as IBM Watson[4] and Microsoft Cognitive Services[5] can extract concepts, emotions, entities, keywords, and relations from unstructured text, and use advanced machine learning algorithms to derive analytics, generate predictions and hypothesis in a scalable way. Therefore, they can offer Cognitive Computing potential to LA, so that the knowledge lost by the existing approaches can be recovered and enriched. This has led to the contribution at the basis of our work.

In this paper, we are interested in supporting the development of LA tools powered by Cognitive Computing by investigating (i) how we can extract and merge features extracted from micro-learning videos to improve their representation in the eyes of machine learning algorithms, and (ii) which machine learning algorithm is best at taking advantage of such features in terms of effectiveness and efficiency in micro-learning video classification. We aim to bridge LA and Cognitive Computing to analyze micro-learning video collections through a fully-automated pipeline, going over traditional term-based methods (e.g. TF-IDF). More precisely, we propose an efficient and effective approach to classify a collection of educational videos on pre-existing categories which uses (i) a novel Speech-to-Text tool to get video transcripts, (ii) cutting-edge Natural Language Processing and Cognitive Computing tools to extract semantic concepts and keywords for their representation, and (iii) Apache Spark as Big Data technology for scalability. Several classifiers have been trained on feature vectors extracted by Cognitive Computing tools on a dataset downloaded from Coursera. Considering the experimental results, our approach promises to improve micro-learning video classification performance and have practical implications in the development of LA tools (e.g. recommendation and automation).

The contribution of this paper is threefold.

- We arranged a large-scale dataset composed by features extracted from 10,328 videos downloaded from Coursera. They are pre-annotated with 7 general-level categories and 34 specific-level categories. The dataset[6] is made publicly available to promote further research in the field.
- We propose a fully-automated approach for effective and efficient classification of micro-learning videos based on Cognitive Computing for feature extraction and on top of Big Data technologies for fast computation, going over traditional term-based methods, with several practical implications in the development of cognitive-driven LA tools.
- We provide an extensive performance evaluation in terms of effectiveness and efficiency for micro-learning classification, comparing our approach with different combinations of feature types and classification algorithms on the collected dataset to assess which one is best at taking advantage of the peculiarities of micro-learning videos.

The rest of the paper is organized as follows. Section 2 outlines the background at the basis of our contribution. Next, Section 3 presents our approach for micro-learning video classification, detailing the components and how they work together. Section 4 describes the experimental evaluation we performed. Finally, Section 5 envisions the practical implications of our approach on LA and Section 6 concludes the paper.

2. Background

In this section, we first provide an overview of how the analysis of learning resources can play a significant role in LA. Then, we focus on the advantage of integrating Cognitive Computing techniques for video content exploration, highlighting similarities and differences between our approach and the existing ones. Finally, we describe relevant Big Data technologies and their essential contribution to achieve fast computation.

2.1. Learning Analytics

Existing LA approaches typically exploit data generated by users during normal interactions with e-learning technologies (Gašević, Dawson, Rogers, & Gasevic, 2016). Several studies focused on the prediction of either the students at risk of failing or the students' grades (Almutairi, Sidiropoulos, & Karypis, 2017; Costa, Fonseca, Santana, de Araújo, & Rego, 2017). However, they tend not to consider how the performance of algorithmic techniques at the basis of LA tools influences the students' behavior during common e-learning tasks. For instance, they investigate whether providing a taxonomy of videos helps students, without considering the performance of the underlying classification algorithm which is essential as well. Moreover, no deep semantic exploration of the resources selected as appropriated for a given learning context is performed. It follows that modern LA tools should consider the content of resources in addition to the interaction with them, especially to find the most appropriate ones.

In the next generation of personalized learning environments, it is essential to provide resources tailored to the learner's need while integrating interactions, skills and competencies with the mapping of knowledge of disciplines (Siemens, 2013). In this direction, powering LA tools with content-based resource analysis promises to support content managers during video organization and learners during search in well-organized educational

[4] https://www.ibm.com/watson/.
[5] https://www.microsoft.com/cognitive-services/en-us.
[6] Please contact the authors via e-mail to get a copy of the dataset.

environments. We point out a novel viewpoint in modeling LA tools, leveraging information conveyed by learning resources.

2.2. Cognitive computing for content analysis

Cognitive computing represents a new appealing model to develop applications capable of working well where traditional methods fail because they are limited by a high level of uncertainty, noise and complexity in data processing. In general, its traditional applications are based on text mining techniques while there are other ones based on data mining, machine learning and deep learning (Bengio, 2009). Cognitive Computing systems can ingest data from external resources, add functions to identify patterns and relationships in large and unstructured data, and consequently interpret massive amount of varieties of them. In this way, they attempt to mimic human behaviors, adding the ability to manage huge amount of data (Chen, Argentinis, & Weber, 2016). Therefore, Cognitive Computing services can reduce the gap between the interpretation and the summarization of information, learning from a large set of interrelated concepts, providing a key for their comprehension and generalization.

One of the most popular Cognitive Computing system is IBM Watson. On February 2011, it was introduced as a question-answering system based on advanced natural language processing, information retrieval, knowledge representation, and automated reasoning. It is currently composed by 14 tools which provide analytics and interaction services. In this set, the Natural Language Understanding analytics tool[7] provides a collection of natural language APIs which enables developers to explore unstructured text, detecting and inferring high-level insights. This service applies multiple technologies to enable the comprehension of vast data resources, independently from their domain. It provides, among others, the following set of functions: entity extraction, sentiment analysis, keyword extraction, concept tagging, relation extraction, language detection, text extraction, microformats parsing, feed detection and linked data. Processing the semantic context provided by these functions helps to infer high-level features characterizing the topic of a text, without suffering from noisy data. In this paper, we use high-level features to indicate features resulting from content exploration carried out by Natural Language Understanding tools.

2.3. Video content analysis

Multimedia classification and indexing are two tasks required to organize and store resources so that they can be quickly retrieved. Several machine learning techniques try to automatize these tasks in an accurate and non time-consuming way.

The term classification is defined as the procedure to select the most appropriate category for a resource from a predefined taxonomy. In recent years, this topic has been widely explored (Farid, Zhang, Rahman, Hossain, & Strachan, 2014; Ren, Peetz, Liang, Van Dolen, & De Rijke, 2014; Song & Roth, 2014; Tang, He, Baggenstoss, & Kay, 2016). Similarly to the previous works, we have adopted some pre-existing classification algorithms such as Support Vector Machine (SVM) (Orazio, Landis, Palmer, & Schrodt, 2014; Tang, Qin, & Liu, 2015; Xing, Wang, Zhang, & Liu, 2014), both in its C4.5 (Quinlan, 2014) and Stochastic Gradient Descent (SGD) variant (Bottou, 2010), Decision Trees (DTs) (Breiman, Friedman, Olshen, & Stone, 1984), and Random Forests (RFs) (Breiman, 2001). They have been selected due to their wide adoption in previous works. Moreover, they have achieved good performance on large-dimensional spaces.

However, it is well-known that the usual characteristics of a video presented in form of visual frames and audio tracks make the classification harder in relation to text content. In order to address this issue, researchers tend to use video metadata (Algur & Bhat, 2016) and video content information such as visual frame sequences (Gkalelis & Mezaris, 2014; Yang & Wang, 2016), audio tracks (Feki, Ammar, & Alimi, 2016), transcripts (Ellis, Jou, & Chang, 2014; Kim et al., 2014), or multiple combinations of them (Chen, Cooper, Joshi, & Girod, 2014; Garibotto et al., 2013; Wang & Ji, 2015). In spite of good results obtained using low-level features, emerging semantic-based alternatives have a huge potential to better describe video content.

Recently, some studies have been focused on higher level features to model the content of videos. In Jiang, Yu, Meng, Mitamura, & Hauptmann, 2015, the authors presented a novel system for content understanding and semantic search on a video collection based on low and high level visual features. In a similar way, Oh et al., 2014 investigated the problem of detecting or classifying single video-clips by means of the event occurring in them, which is defined by a complex collection of elements including people, objects, actions, and relations among them. The authors in Dalton, Allan, & Mirajkar, 2013 proposed to generate queries on videos exploiting Automatic Speech Recognition (ASR) and Optical Character Recognition (OCR) directly. However, we point out that the type of video represents a relevant characteristic to better understand their content. Hence, the information channels used for a reliable analysis should be selected in relation to the way the information is mainly transmitted. In micro-learning videos, the greatest amount of knowledge is transmitted by voice. Therefore, we have chosen video transcripts as source of information. Similar approaches have been already adopted, considering the extraction of transcripts from videos through ASR systems and the detection of features from them. For instance, the authors in Ellis et al., 2014 used the transcripts to improve a sentiment classification of general videos. Besides, Kim et al., 2014 proposed various video navigation techniques in educational context. One of their approaches used keyword clouds computed from the transcripts using TF-IDF scores, and provided a topic summarization to help learners to understand the video content. In the e-learning domain, retrieval and classification tasks were performed by Basu et al., 2016 using Latent Semantic Allocation to find and represent topics conveyed by a video lecture. The method demonstrated the ability of better representing in comparison to the information extracted from title and tags associated to a video. Other existing works tend to consider external contents or corpus for performing classification tasks. For instance, the authors in Chen, Cao, Song, Zhang, & Li, 2010 used Wikipedia as an open resource knowledge to match videos with the most suitable category based on a taxonomy. Differently, we used APIs making our approach independent on updates and exploitable on new content with no remarkable changes.

The analysis of audio tracks as textual transcripts is an attractive way to be exploited as the most rich channel of information for micro-learning videos. In this direction, we investigate how novel feature types extracted from transcripts by Cognitive Computing tools can improve their classification. In contrast to the previous works, we face video content analysis exploiting the Natural Language Understanding capabilities provided by IBM Watson, enriching TF-IDF results with semantic information. Embedding the service in our approach, we then focus on the analysis of various feature sets and classification algorithms to achieve better classification performance.

[7] https://www.ibm.com/watson/developercloud/natural-language-understanding.html.

2.4. Big Data technologies

The Big Data movement consists of increasingly powerful and relatively inexpensive computing platforms with fault-tolerant storage and processing carried out through thousands of processors. The current volume of data managed by existing systems has surpassed the processing capacity of the traditional ones and this applies to data mining as well (Wu, Zhu, Wu, & Ding, 2014). The arising of new technologies and services (e.g. Cloud Computing and Grid Computing) and the reduction in hardware price have led to an ever-growing rate of information on the Internet. Consequently, Big Data applications in various domains, such as economics and finance (Aymerich, Fenu, & Surcis, 2009) and computer networks (Fenu & Nitti, 2011), can be efficiently moved into the cloud to analyze larger amounts of data.

In last years, several platforms for large-scale processing have tried to face the problem of Big Data (Fernández et al., 2014). These platforms try to bring closer distributed technologies to standard users (e.g. engineers and data scientists) by hiding the technical nuances derived from distributed environments. On the other hand, Big Data platforms also require additional algorithms that give support to relevant tasks, like big data preprocessing and analytics. Standard algorithms for those tasks must be also redesigned if we want to learn from large-scale datasets. It is not a trivial thing and presents a big challenge for researchers. The first framework that enabled the processing of large-scale datasets has been Map Reduce. This tool has been intended to process and generate huge datasets in an automatic and distributed way. By implementing two primitives, Map and Reduce, the user is able to use a scalable and distributed tool without worrying about technical nuances such as failure recovery, data partitioning or job communication. Apache Hadoop[8] has emerged as the most popular open-source implementation of Map Reduce, maintaining the aforementioned features. In spite of its great popularity, Map Reduce is not designed to scale well when dealing with iterative and online processes typical in machine learning and stream analytics. Apache Spark[9] has been designed as an alternative to Hadoop capable of performing faster distributed computing by using inmemory primitives. Spark is built on top of a new abstraction model called Resilient Distributed Datasets (RDDs). This versatile model can control the persistence and the partition of data effectively and efficiently.

3. The proposed approach

In this section, we describe the approach for micro-learning video classification proposed in this paper. Fig. 1 depicts its components and how they work together.

The Classification Manager orchestrates the overall process. First, it calls the Pre-Processing Module to extract the transcripts from the videos included into the dataset. Then, these transcripts are handled by the Feature Extraction Module which computes the corresponding features. During training, the Classification Manager sends both features and category labels to the Classifier Module. During testing, only the features are sent, and the returned categories are matched with the original ones stored in the dataset to evaluate the performance. We detail all the modules in the following subsections.

3.1. Big Data manager

We have employed Apache Spark 1.6.1 and MLlib library. MLlib is the Spark's machine learning library aimed at making practical machine learning scalable and easy. It includes common learning algorithms and utilities, such as classification, regression, clustering, collaborative filtering, dimensionality reduction, as well as lower-level optimization primitives and higher-level pipeline APIs. Including it makes our approach general and flexible. We can easily use any classifier available within the MLlib library, whose code is already optimized with the Map-Reduce paradigm to run over a cluster. We might also use any other classifier of other libraries not specific for Apache Spark. The cluster has been preliminarily configured to use over 100 machines within the laboratory of our department.

3.2. Pre-Processing module

This module takes a micro-learning video as an input and returns the cleaned version of the associated transcript as an output. Given a micro-learning video v, the module sends it to the Speech-to-Text service[10] of the IBM Watson's suite, which combines grammar rules with knowledge of audio signals for translating the spoken language of an audio track in its written form. Its choice depends on the fact that it is one of the most accurate and easily manageable speech-to-text tools (Marković, Popat, Antonacci, Sarti, & Kumar, 2016). The service returns a plain text $t(v)$ corresponding to the transcript of v. The module converts all the words in $t(v)$ to lowercase and each one of them is compared with the WordNet[11] dictionary in order to get the correct one, w', if w contains errors; otherwise, w and w' have the same value. Each word w in $t(v)$ is substituted in $t(v)$ with the correct w', obtaining $t'(v)$. Finally, the module removes stop-words from $t'(v)$ and returns it as cleaned transcript.

3.3. Features extraction module

The Features Extraction Module has the purpose of representing a micro-learning video with a set of features. It takes a cleaned transcript from the Pre-Processing Module as input and returns a set of pairs where the first element is the identifier string of the feature and the second element is the relevance of that feature for the corresponding transcript. The relevance value spans in the range [0,1] where a value closer to 0 represents a low relevance and a value closer to 1 represents a high relevance of the corresponding feature for the transcript.

From the transcript, the module can extract TF-IDF features and high-level features. It returns one of these feature sets according to a flag set up into the Feature Selector submodule. More precisely, for the high-level features, the module calls the Natural Language Understanding API provided by the IBM Watson's suite, which contains a collection of natural language processing functions aiming at extracting keywords, concepts, entities and so on from texts. Currently, the module exploits only concepts and keywords, but it can be easily extended to a wider set of features. This service works well for semantic content extraction since it contains a wide range of pre-trained models, and can process large samples of actual human language to derive rules governing the natural language in a sentence. In our case, the Features Extraction Module shapes keywords and concepts into vectors. The first one includes important topics typically used when indexing data. The service

[8] http://hadoop.apache.org/.
[9] http://spark.apache.org.
[10] https://www.ibm.com/watson/developercloud/speech-to-text.html.
[11] https://wordnet.princeton.edu/.

Please cite this article in press as: Dessì, D., et al., Bridging learning analytics and Cognitive Computing for Big Data classification in micro-learning video collections, Computers in Human Behavior (2018), https://doi.org/10.1016/j.chb.2018.03.004

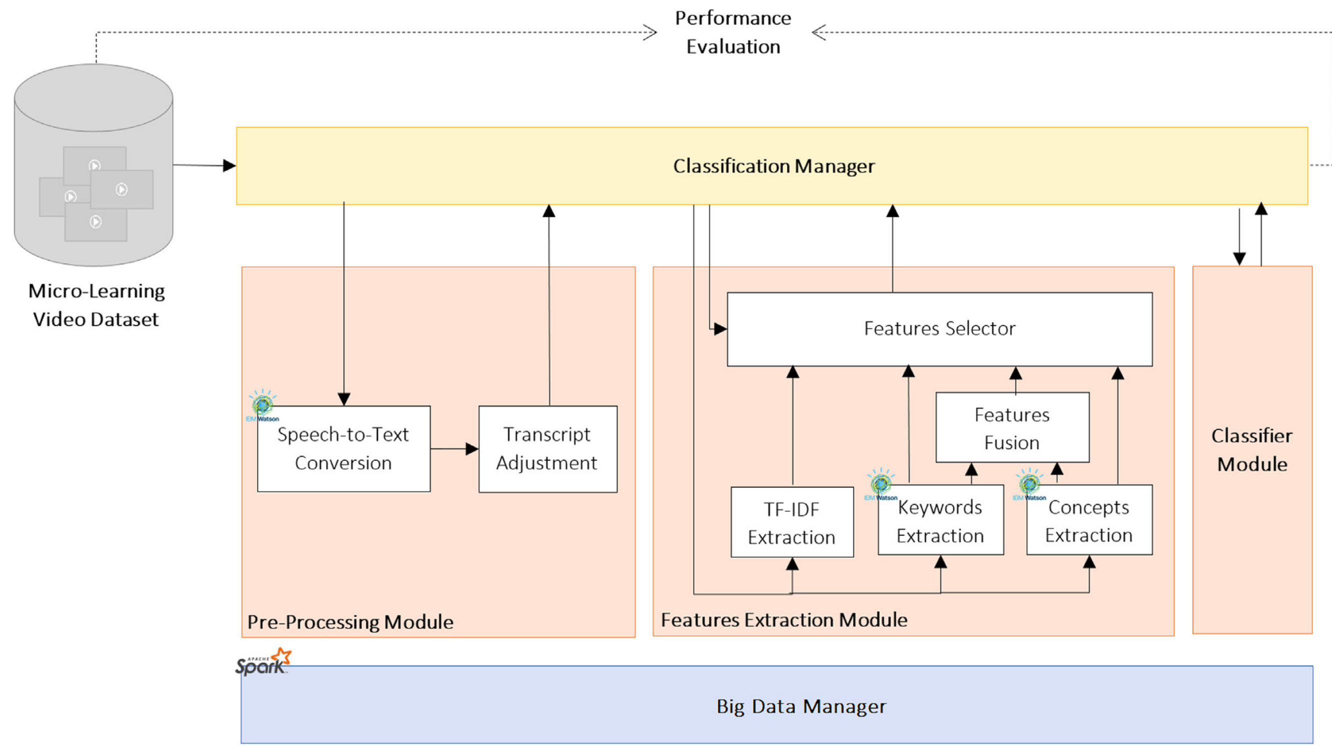

Fig. 1. A schema for the proposed approach for micro-learning video classification and how the components work.

identifies and ranks keywords directly mentioned in the text. The second one represents concepts not necessarily referenced in the text, but with which the transcript text can be associated. These concepts are inferred from the relations between syntactic and semantic elements of the sentences. Both for concepts and keywords, the service computes a weight that indicates the relevance we exploited in our vectors. These feature sets enable our approach to perform high-level analysis than just TF-IDF.

Given a cleaned transcript $t'(v)$ as returned from the Pre-Processing Module, the module first computes for each word $w \in t'(v)$ its TF-IDF value building the TF-IDF vector $tf\text{-}idf(t'(v))$. Then, it sends $t'(v)$ to the Natural Language Understanding service and requests to obtain a concepts vector $c(t'(v))$ and a keywords vector $k(t'(v))$. The service returns them as JSON data. For each vector, a list of pairs is returned where the first element of each pair is the string identifier and the second element is the relevance associated to it in the range [0,1]. After that, the module builds a unique feature vector $kc(t'(v))$ by concatenating $c(t'(v))$ and $k(t'(v))$. As an example, we consider a short segment of a video transcript about computer networks. The text and the corresponding concepts and keywords extracted using IBM Watson are shown in Fig. 2. "Computer network" is the only concept extracted even though it is not directly mentioned in the text. A number of four keywords were returned. In the case of a collision between two identifiers from concepts and keywords vectors, the module computes the mean between the associated relevance values and stores a single instance of the identifier accordingly.

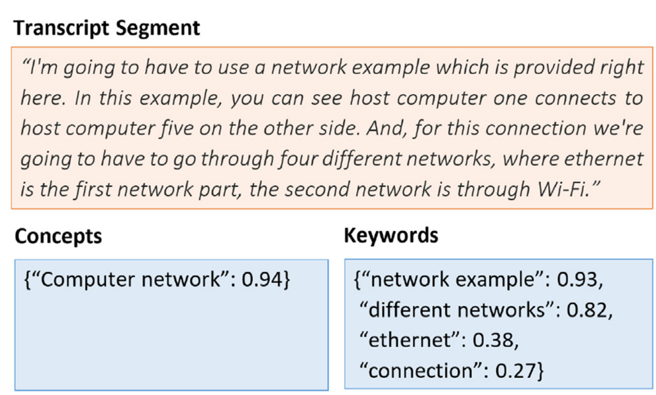

Fig. 2. Example of extraction of features with IBM Watson.

3.4. Classifier module

The Classifier Module aims at finding the most appropriate category for a given video using the underlying classifier trained on a number of labeled samples. The classifier implements a function $f(t'(v)) \rightarrow c$ where $f(t'(v))$ is the features vector of $t'(v)$ and c is the category returned by the classifier and of the given taxonomy. The module can implement any classification algorithm, independently from the feature type. In particular, our approach tests DT, SVM, RF, and SVM+SGD since they are the most widely used algorithms as emerged from literature review. However, our approach does not depend on this design choice and any classification algorithm can be used.

4. Experimental study

In this section, we describe the experiments we performed to explore both the effectiveness and the efficiency of our approach. First, we describe how we generated the dataset and present the evaluation measures we adopted. Then, we present the performance of the proposed approach using several combinations of different feature sets and classifiers.

Please cite this article in press as: Dessì, D., et al., Bridging learning analytics and Cognitive Computing for Big Data classification in micro-learning video collections, Computers in Human Behavior (2018), https://doi.org/10.1016/j.chb.2018.03.004

4.1. Dataset description

We collected one real-world micro-learning video dataset from Coursera to validate our methodology. This task was necessary since, to the best of our knowledge, there exists no benchmark dataset for micro-learning video classification. We collected 10,328 videos from 617 courses whose primary language is English. Coursera pre-assigned the courses to 7 general-level categories and 34 specific-level categories. Coursera has a courses catalog; each course has one general-level category and one specific-level category. Moreover, each course contains a set of videos. Each downloaded video was assigned to the same categories of the course it belongs.

Each extracted video is assigned to one general-level category and one specific-level category and contains a number of words ranging from 200 to 10.000 (avg. 1525; std. 1017). The overall distribution of videos per category is reported in Fig. 3 and Fig. 4. The dataset is challenging, since it contains fine-grained categories that require subtle details to differentiate (e.g. Business Essentials and Business Strategy). Moreover, video transcripts contain less words than documents typically used in text classification. The language style is different, since the transcripts are derived from speaking activities, not written.

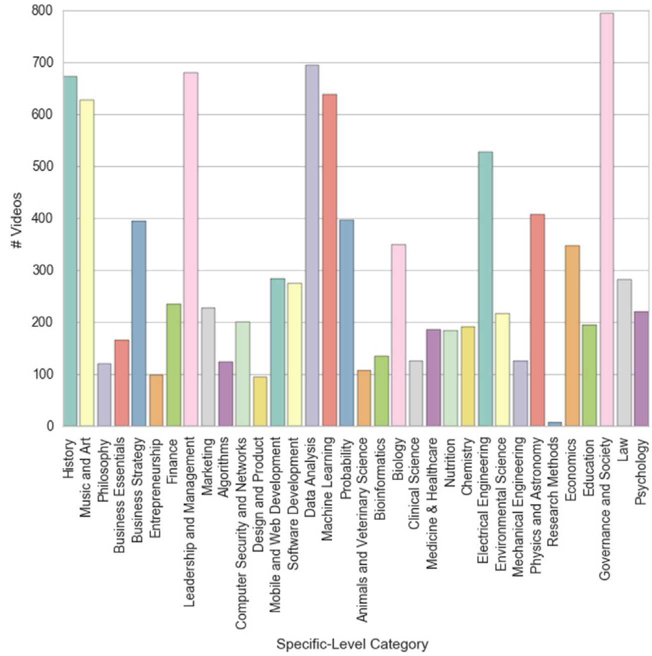

Fig. 4. Video distribution over specific categories in our extracted dataset.

4.2. Performance measures

We mapped the video classification problem as a text classification problem. Hence, we adopted the most common performance measures for this task: precision, recall, and F-measure. Because we were dealing with a multi-class classification problem, we first define how to calculate them for each category.

Given a specific category c_i from the category space $c_1, \ldots, c_n$, the corresponding precision P_{c_i}, recall R_{c_i}, and F-measure $F_{1_{c_i}}$ are defined by the following formulas:

$$P_{c_i} = \frac{TP_{c_i}}{TP_{c_i} + FP_{c_i}} \quad R_{c_i} = \frac{TP_{c_i}}{TP_{c_i} + FN_{c_i}} \quad F_{1c_i} = 2\frac{R_{c_i} \cdot P_{c_i}}{R_{c_i} + P_{c_i}}$$

where TP_{c_i} (true positives) is the number of videos assigned correctly to the category c_i, FP_{c_i} (false positives) is the number of videos that do not belong to the category c_i, but they are assigned to this category incorrectly and FN_{c_i} (false negatives) is the number of videos that actually belong to the category c_i, but they are not assigned to this class.

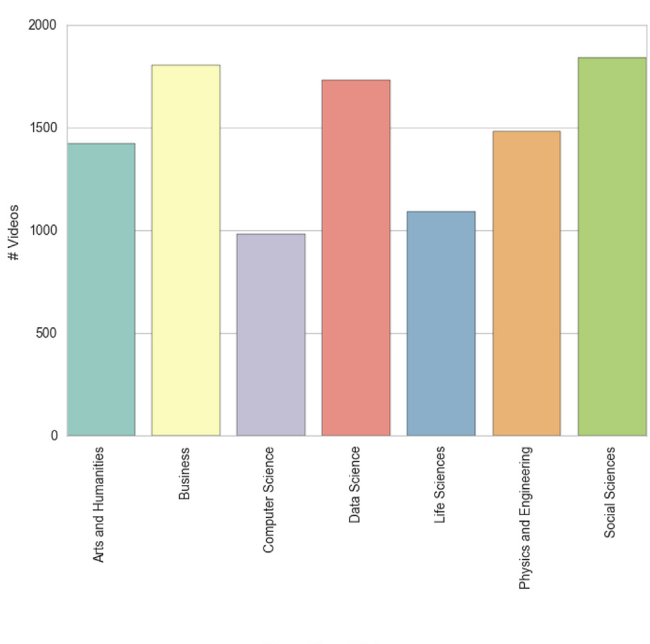

Fig. 3. Video distribution over general categories in our extracted dataset.

Then, we need to compute the average performance of a binary classifier (i.e. one for each category) over multiple categories. There are three main methodologies for the computation of averaged metrics (see Sokolova & Lapalme, 2009 for more information). They can be summarized as follows:

- *Micro-Averaged* (*Micro*). The metrics are globally calculated by counting the total number of true positives, false negatives and false positives, with no category differences.
- *Macro-Averaged* (*Macro*). The metrics are locally calculated for each category and the unweighted mean is computed. This does not consider categories imbalance.
- *Weighted-Averaged* (*Weight*). The metrics are locally calculated for each category, then their average is obtained by weighting each category metric with the number of instances of the category in the dataset. Therefore, each category does not contribute equally to the final average and some of them contribute more than the others.

To test the performances of our approach, we considered nine scores obtained from the combination of the metrics (precision, recall, F-measure) with the modalities for average computation (micro, macro, weighted), for each one of the two categories spaces of our dataset (general-level and specific-level).

Please cite this article in press as: Dessì, D., et al., Bridging learning analytics and Cognitive Computing for Big Data classification in micro-learning video collections, Computers in Human Behavior (2018), https://doi.org/10.1016/j.chb.2018.03.004

4.3. Performance evaluation

In this section, we focus on exploring the effectiveness and the efficiency of our approach for a certain number of combinations of the features sets, classification algorithms and metrics. The software was developed using PyCharm 2016.2 environment. We used Python as programming language. For each experiment, we adopted a 10-fold stratified cross validation, so that the folds were made by preserving the percentage of samples for each category, then the reported results were averaged over ten runs. Apache Spark was set underneath and the code was developed on top using map reduce functions that allows scalability in the cluster of more than 100 machines we configured. For the experiments, 12 cores were previously allocated.

Feature-Classifier Combinations In order to evaluate our approach, we tested it with a number of alternative combinations of four features representations and four classification algorithms. The features representations included TF-IDF (baseline), concepts, keywords, and the combination of keywords and concepts. While, the classifiers were the implementations of the following algorithms provided by Buitinck et al., 2013:

- Decision Tree (DT). This method creates a model based on C4.5 algorithm, predicting the value of a target variable by learning decision rules from data features. The model has a root node containing all data features of the training set. Then, the root node is split into several children according to a given criteria. This process recursively continues on children until there are not other nodes to be split.
- Support Vector Machine (SVM). This method plots each data item as a point in n-dimensional space (where n is number of features) with the value of each feature being the value of a particular coordinate. Then, it finds the set of hyper-planes that better differentiate the categories. A linear combination of vectors determines the location of the decision boundaries producing the best separation of categories.
- Random Forest (RF). This method is a meta estimator that fits a number of decision tree classifiers on various random sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. Each decision tree, individually, is a weak classifier, while all the decision trees taken together would be a strong classifier.
- Support Vector Machine + Stochastic Gradient Descent (SVM+SGD). This method extends the standard SVM implementation including the SGD algorithm during training. SGD finds the best coefficients describing the decision boundaries through a classification function which minimizes a hinge loss function and allows performing training over large data while reducing the computation time.

Computational Time Evaluation We evaluated the efficiency of the proposed approach in terms of the size of the features vectors used and the time required to perform both the training and the test for a given classifier.

Table 1 summarizes the basic statistics with respect to the different feature sets. The first column indicates the name of the feature set, the second column shows its size, while the other two columns report the average and the standard deviation of the number of non-zero elements. The vector sizes for the combination concepts and keywords is greater than all the others. However, the average number of its non-zero elements is smaller than that of the TF-IDF. Hence, high-level features can be more discriminative. Table 2 reports the total time required for a given algorithm with a given features set to perform the training and the test phases over a single fold. In the results, SVM+SGD algorithm provided the best computational time, especially in case of concepts with no large vector size. The computational time mostly depends on the particular classifier and the number of features. For example, the computational time for SVM+SGD classifier using concepts as features takes 0.05 s due to the lower time required by SVM+SGD and the lower size of the concepts feature set.

Precision-Recall Analysis We evaluated the performances of all the classifiers trained on TF-IDF, keywords, concepts, and keywords plus concepts using precision, recall and F-measure. The results in Table 3 indicate that using keywords outperforms TF-IDF only with the SGD approach. Using concepts generally gives better results than using keywords, and as with the use of keywords, they outperform TF-IDF in case of SGD algorithm. With concepts, we obtain higher precision and lower recall, particularly for macro and weighted evaluations. For SVM and SVM+SGD, keywords plus concepts outperforms TF-IDF.

Table 4 shows the results of the classifiers on specific categories. In this case, SGD classifier trained using keywords continues to outperform TF-IDF, while its performance decreases if trained on concepts. When combining keywords and concepts, the overall performance improves and, in case of SVM+SGD, can be considered better than TF-IDF with an improvement up to 9%. The worst case shows a maximum loss of 7% when SVM is adopted. As far as the specific categories classification is concerned, it is worth noticing that they are not equally distributed as the general categories (as clearly shown in Tables 3 and 4) and this high variance negatively affects the classification using high-level features (keywords, concepts and their combination), especially those with a low number of features for the training.

Overall Evaluation In most cases, our approach can produce good performance, regardless the increasing algorithm complexity in terms of video size or transcript size. In fact, the first one

Table 1
Basic statistics about the used features sets.

Feature Set	Total Size	Avg.	Std. Dev.
TF-IDF (baseline)	117,073	260	513
Concepts	29,952	21	13
Keywords	332,023	78	26
Keywords + Concepts	361,975	99	31

Table 2
Computational time for both training and test steps. For each pair of features set and category level, bold values represent the computational time spent by the fastest algorithm.

Features Set	Algorithm	Execution Time [s]	
		General Categories	Specific Categories
TF-IDF (baseline)	DT	5.23	7.92
	RF	12.20	39.81
	SVM	3.70	16.25
	SVM+SGD	**0.25**	**1.12**
Keywords	DT	3.32	7.21
	RF	11.06	43.96
	SVM	3.32	11.21
	SVM+SGD	**0.20**	**1.00**
Concepts	DT	1.13	1.88
	RF	2.38	7.67
	SVM	0.31	2.21
	SVM+SGD	**0.05**	**0.20**
Keywords + Concepts	DT	4.00	8.34
	RF	13.53	50.38
	SVM	3.95	12.84
	SVM+SGD	**0.26**	**1.06**

Please cite this article in press as: Dessì, D., et al., Bridging learning analytics and Cognitive Computing for Big Data classification in micro-learning video collections, Computers in Human Behavior (2018), https://doi.org/10.1016/j.chb.2018.03.004

Table 3
Performance measures for micro-learning video classification over general categories.

Feature Set	Algorithm	Precision			Recall			F-Measure		
		Micro	Macro	Weight	Micro	Macro	Weight	Micro	Macro	Weight
TF-IDF (baseline)	DT	0.48	0.56	0.55	0.48	0.47	0.48	0.48	0.48	0.48
	RF	0.58	0.61	0.60	0.58	0.57	0.58	0.58	0.57	0.57
	SVM	**0.70**	**0.72**	**0.71**	**0.70**	**0.70**	**0.70**	**0.70**	**0.70**	**0.69**
	SVM+SGD	0.62	0.62	0.62	0.62	0.62	0.62	0.62	0.61	0.61
Keywords	DT	0.35	0.49	0.47	0.35	0.32	0.35	0.35	0.32	0.33
	RF	0.43	0.51	0.49	0.43	0.41	0.43	0.43	0.40	0.41
	SVM	0.65	0.68	0.67	0.65	0.65	0.65	0.65	0.65	0.65
	SVM+SGD	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.65	0.65
Concepts	DT	0.44	0.68	0.66	0.44	0.42	0.44	0.44	0.44	0.45
	RF	0.52	0.67	0.66	0.52	0.51	0.52	0.52	0.53	0.53
	SVM	0.62	0.64	0.63	0.62	0.62	0.62	0.62	0.62	0.61
	SVM+SGD	0.63	0.64	0.64	0.63	0.63	0.63	0.63	0.63	0.62
Keywords + Concepts	DT	0.46	0.65	0.64	0.46	0.44	0.46	0.46	0.47	0.47
	RF	0.55	0.64	0.63	0.55	0.54	0.55	0.55	0.55	0.55
	SVM	0.69	0.71	0.70	0.69	0.69	0.69	0.69	0.69	**0.69**
	SVM+SGD	0.68	0.69	0.69	0.68	0.68	0.68	0.68	0.67	0.67

Table 4
Performance measures for micro-learning video classification over specific categories.

Feature Set	Algorithm	Precision			Recall			F-Measure		
		Micro	Macro	Weight	Micro	Macro	Weight	Micro	Macro	Weight
TF-IDF (baseline)	DT	0.46	0.53	0.52	0.46	0.42	0.46	0.46	0.44	0.46
	RF	0.58	0.59	0.59	0.58	0.54	0.58	0.58	0.54	0.56
	SVM	**0.71**	**0.73**	**0.73**	**0.71**	**0.67**	**0.71**	**0.71**	**0.67**	**0.70**
	SVM+SGD	0.58	0.64	0.69	0.58	0.57	0.58	0.58	0.57	0.61
Keywords	DT	0.34	0.50	0.48	0.34	0.32	0.34	0.33	0.36	0.34
	RF	0.43	0.49	0.49	0.43	0.41	0.43	0.43	0.42	0.42
	SVM	0.62	0.70	0.67	0.62	0.55	0.62	0.62	0.57	0.61
	SVM+SGD	0.66	0.66	0.70	0.66	0.65	0.66	0.66	0.63	0.66
Concepts	DT	0.39	0.49	0.52	0.39	0.38	0.39	0.39	0.39	0.40
	RF	0.48	0.50	0.53	0.48	0.46	0.48	0.48	0.45	0.47
	SVM	0.57	0.58	0.59	0.57	0.53	0.57	0.57	0.53	0.56
	SVM+SGD	0.57	0.55	0.58	0.57	0.54	0.57	0.57	0.53	0.56
Keywords + Concepts	DT	0.41	0.52	0.52	0.41	0.40	0.41	0.41	0.42	0.41
	RF	0.51	0.54	0.54	0.51	0.50	0.51	0.51	0.49	0.50
	SVM	0.66	0.70	0.69	0.66	0.60	0.66	0.66	0.62	0.65
	SVM+SGD	0.67	0.64	0.69	0.66	0.65	0.66	0.67	0.62	0.66

influences only the time required to extract transcripts from videos. No assumptions can be done on transcript sizes in relation to video sizes since the number of words included into a transcript depends on the amount of content orally provided by the teacher during the video lesson. The second one has an impact on feature extraction. Training and test are not directly influenced by the transcript size because the size of each feature vector depends on the size of the word dictionary used to map features. Moreover, the number of relevant features extracted from a long transcript can be less than those extracted from a short one due to repetitions and stop words.

Considering the trade-off between effectiveness in terms of precision, recall, F-measure and efficiency in terms of computational time, the combinations achieving best results include SVM or SVM+SGD as algorithm and TF-IDF or Keywords + Concepts as features. However, SVM+SGD algorithm is generally over ten times faster than SVM. With SVM+SGD, our approach using Keywords + Concepts outperforms that using TF-IDF from 7% to 9% in terms of precision, recall and F-measure. The combination using SVM+SGD and Keywords + Concepts achieves performance comparable with that using SVM and TF-IDF in terms of effectiveness, but strongly better in efficiency. The experimental results demonstrate that Keywords + Concepts combined with SVM+SGD can scale well, maintaining good performances in almost all cases.

5. Practical implications

The promising experimental results inspire several practical applications of our approach worthy of future study to develop cognitive-driven LA tools within a broad domain of educational research. This section presents a set of relevant examples.

One of the primary application domains of LA and data mining techniques is the recommendation of learning materials to students. Semantic techniques can lead to the evolution from a TF-IDF-based to a concept-based representation of items and user profiles. In this context, content-based recommendation techniques can combine our classification approach with students data in order to build tailored student's profiles from their set of watched videos, by looking for most similar resources to the ones that a student has previously used. Using deep semantic content analytics, students can gain improvements in learning because they can be directly driven to resources that best fit their interests, reducing the time required for retrieval.

Since LA makes use of different data mining and machine learning techniques for analysis of educational content, our approach has a potential for automating repetitive and time-consuming tasks usually performed by educational researchers and practitioners. These include automation of micro-learning classification processes, tagging generation and organization of

Please cite this article in press as: Dessì, D., et al., Bridging learning analytics and Cognitive Computing for Big Data classification in micro-learning video collections, Computers in Human Behavior (2018), https://doi.org/10.1016/j.chb.2018.03.004

learning resources. MOOC platforms usually contain a large amount of videos which need to be organized based on their content for being provided to students. Content managers need to know the concepts behind videos which should be accurately allocated into a taxonomy for a fast and qualitative organization. With our approach, micro-learning videos can be automatically placed into a predefined taxonomy, increasing the quality of their arrangement into MOOC platforms while reducing the effort required to content managers.

Another primary challenge of working with large amounts of video content is indexing and retrieval. The increasing development of advanced multimedia applications requires new technologies for organizing and retrieving from content databases of digital videos. To this aim, video content must be described and adequately coded. Our approach can be easily extended to index and retrieval, allowing semantic annotation and querying in video databases. In this way, learners can be supported during the search of the most appropriate content fitting their requirements. Even more, our automated pipeline facilitates the development of video-to-video search tools, making a step forward to mitigate the well-known semantic gap problem.

Since the majority of learning forms involve use and creation of videos, several analytics techniques have been applied to investigate them. In an e-learning platform, the success of LA tools requires both a simple user interface to directly interact with users and algorithms supporting them for achieving and managing resources they are interested in. Our contribution provides a back-end powered by Cognitive Computing and Big Data for several LA tool interfaces which require to work with micro-learning videos.

6. Conclusions

We have presented a novel method for micro-learning video classification. Our approach extracts transcripts from videos using novel speech-to-text systems and exploits Cognitive Computing to represent the features of each video transcript. Moreover, it leverages Big Data technologies for fast computation.

The experimental results show how our approach achieves good performance in most cases for both computational time and precision-recall analysis. Our feature representation combines concepts and keywords extracted from cutting-edge Cognitive Computing tools and exploits the semantic behind the text that traditional approaches fail to capture. Considering the experimental results, we expect our approach powered by Cognitive Computing can facilitate the development of cognitive-driven LA tools aimed at supporting content managers to arrange micro-learning video collections while improving how learners can exploit categories for exploring it.

In next steps, we plan to test our approach on larger datasets of micro-learning videos, even collected from YouTube and annotated by using crowdsourcing services (e.g. Crowdflower[12]). Leveraging Apache Spark allows the method to be scalable enough to process and classify millions of videos. Moreover, we would embed our approach as a base for LA tools to conduct usability and subjective evaluation as well as quantifying the impact that content-based LA tools have. Furthermore, our approach might be combined with experiential data and interactions of students for the purpose of video recommendations that better fit learners' interests.

With the ever-growing interest in LA as well as the rapid development of Cognitive Computing and Big Data technologies, combining them has a good potential to advance innovation on personalized learning environments. LA services powered by Cognitive Computing promise to shape the future of higher education. Our contribution tries to make a jump-start towards this upcoming era of cognitive-driven education.

Acknowledgments

Mirko Marras and Danilo Dessì gratefully acknowledge Sardinia Regional Government for the financial support of their PhD scholarship (P.O.R. Sardegna F.S.E. Operational Programme of the Autonomous Region of Sardinia, European Social Fund 2014–2020 - Axis III "Education and Training", Thematic Goal 10, Priority of Investment 10ii, Specific Goal 10.5).

The authors gratefully acknowledge for the financial support: Sardinia Regional Government (Convenzione triennale tra la Fondazione di Sardegna e gli Atenei Sardi Regione Sardegna L.R. 7/2007 annualitá 2016 DGR 28/21 del 17.05.2017, CUP: F72F16003030002) and Italian Ministry of Education, University and Research - Program Smart Cities and Communities and Social Innovation project ILEARNTV (D.D. n.1937 del 05.06.2014, CUP F74G14000200008 F19G14000910008).

References

Algur, S. P., & Bhat, P. (2016). Web video mining: Metadata predictive analysis using classification techniques. *International Journal of Information Technology and Computer Science, 8*(2), 69.

Almutairi, F. M., Sidiropoulos, N. D., & Karypis, G. (2017). Context-aware recommendation-based learning analytics using tensor and coupled matrix factorization. *IEEE Journal of Selected Topics in Signal Processing, 11*(5), 729–741.

Aymerich, F. M., Fenu, G., & Surcis, S. (2009). A real time financial system based on grid and cloud computing. In *Proceedings of the 2009 ACM symposium on applied computing* (pp. 1219–1220). ACM.

Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In *Learning analytics* (pp. 61–75). Springer.

Barak, M., Watted, A., & Haick, H. (2016). Motivation to learn in massive open online courses: Examining aspects of language and social engagement. *Computers & Education, 94*, 49–60.

Basu, S., Yu, Y., & Zimmermann, R. (2016). Fuzzy clustering of lecture videos based on topic modeling. In *Content-based multimedia indexing (CBMI), 2016 14th international workshop* (pp. 1–6). IEEE.

Bengio, Y. (2009). Learning deep architectures for ai. *Foundations and trends in Machine Learning, 2*(1), 1–127.

Bottou, L. (2010). Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010* (pp. 177–186). Springer.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees*. Monterey, CA: Wadsworth and Brooks.

Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., et al. (2013). API design for machine learning software: Experiences from the scikit-learn project. In *ECML PKDD Workshop: Languages for data mining and machine learning* (pp. 108–122).

Chen, Y., Argentinis, J. E., & Weber, G. (2016). Ibm watson: How cognitive computing can be applied to big data challenges in life sciences research. *Clinical Therapeutics, 38*(4), 688–701.

Chen, Z., Cao, J., Song, Y., Zhang, Y., & Li, J. (2010). Web video categorization based on wikipedia categories and content-duplicated open resources. In *Proceedings of the 18th ACM international conference on multimedia, MM '10* (pp. 1107–1110). New York, NY, USA: ACM.

Chen, H., Cooper, M., Joshi, D., & Girod, B. (2014). Multi-modal language models for lecture video retrieval. In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 1081–1084). ACM.

Coakley, D., Garvey, R., & ONeill, Í. (2017). Micro-learningadopting digital pedagogies to facilitate technology-enhanced teaching and learning for cpd. In *Empowering 21st Century learners through holistic and enterprising learning* (pp. 237–242). Springer.

Costa, E. B., Fonseca, B., Santana, M. A., de Araújo, F. F., & Rego, J. (2017). Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. *Computers in Human Behavior, 73*, 247–256.

Dalton, J., Allan, J., & Mirajkar, P. (2013). Zero-shot video retrieval using content and concepts. In *Proceedings of the 22nd ACM international conference on information & knowledge management* (pp. 1857–1860). ACM.

Ellis, J. G., Jou, B., & Chang, S.-F. (2014). Why we watch the news: A dataset for exploring sentiment in broadcast video news. In *In proceedings of the 16th international conference on multimodal interaction* (pp. 104–111). ACM.

Farid, D. M., Zhang, L., Rahman, C. M., Hossain, M. A., & Strachan, R. (2014). Hybrid decision tree and naïve bayes classifiers for multi-class classification tasks. *Expert Systems with Applications, 41*(4), 1937–1946.

[12] https://www.crowdflower.com/.

Feki, I., Ammar, A. B., & Alimi, A. M. (2016). Automatic environmental sound concepts discovery for video retrieval. *International Journal of Multimedia Information Retrieval, 5*(2), 105–115.

Fenu, G., & Nitti, M. (2011). Strategies to carry and forward packets in vanet. *Digital Information and Communication Technology and Its Applications*, 662–674.

Fernández, A., del Río, S., López, V., Bawakid, A., del Jesus, M. J., et al. (2014). Big data with cloud computing: An insight on the computing environment, mapreduce, and programming frameworks. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 4*(5), 380–409.

Garibotto, G., Murrieri, P., Capra, A., De Muro, S., Petillo, U., Flammini, F., et al. (2013). White paper on industrial applications of computer vision and pattern recognition. In *International conference on image analysis and processing* (pp. 721–730). Springer.

Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education, 28*, 68–84.

Gkalelis, N., & Mezaris, V. (2014). Video event detection using generalized subclass discriminant analysis and linear support vector machines. In *Proceedings of international conference on multimedia retrieval* (p. 25). ACM.

Jiang, L., Yu, S.-I., Meng, D., Mitamura, T., & Hauptmann, A. G. (2015). Bridging the ultimate semantic gap: A semantic search engine for internet videos. In *Proceedings of the 5th ACM on international conference on multimedia retrieval* (pp. 27–34). ACM.

Kim, J., Guo, P. J., Cai, C. J., Li, S.-W. D., Gajos, K. Z., & Miller, R. C. (2014). Data-driven interaction techniques for improving navigation of educational videos. In *Proceedings of the 27th annual ACM symposium on User interface software and technology* (pp. 563–572). ACM.

Kloos, C. D., Jermann, P., Pérez-Sanagustín, M., Seaton, D. T., & White, S. (2017). *Digital Education: Out to the world and back to the campus*. Springer.

Leach, M., & Hadi, S. M. (2017). Supporting, categorising and visualising diverse learner behaviour on moocs with modular design and micro-learning. *Journal of Computing in Higher Education, 29*(1), 147–159.

Marković, D., Popat, J., Antonacci, F., Sarti, A., & Kumar, T. K. (2016). An informed separation algorithm based on sound field mapping for speech recognition systems. In *Acoustic signal enhancement (IWAENC), 2016 IEEE international workshop* (pp. 1–5). IEEE.

Oh, S., McCloskey, S., Kim, I., Vahdat, A., Cannons, K. J., Hajimirsadeghi, H., et al. (2014). Multimedia event detection with multimodal feature fusion and temporal concept localization. *Machine Vision and Applications, 25*(1), 49–69.

Orazio, V., Landis, S. T., Palmer, G., & Schrodt, P. (2014). Separating the wheat from the chaff: Applications of automated document classification using support vector machines. *Political Analysis, 22*(2).

Quinlan, J. R. (2014). *C4. 5: Programs for machine learning*. Elsevier.

Ren, Z., Peetz, M.-H., Liang, S., Van Dolen, W., & De Rijke, M. (2014). Hierarchical multi-label classification of social text streams. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval* (pp. 213–222). ACM.

Ristoski, P., & Paulheim, H. (2016). Semantic web in data mining and knowledge discovery: A comprehensive survey. *Web Semantics: Science, Services and Agents on the World Wide Web, 36*, 1–22.

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist, 57*(10), 1380–1400.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management, 45*(4), 427–437.

Song, Y., & Roth, D. (2014). On dataless hierarchical text classification. In *AAAI* (Vol. 7).

Sun, G., Cui, T., Guo, W., Beydoun, G., Xu, D., & Shen, J. (2015). Micro learning adaptation in mooc: A software as a service and a personalized learner model. In *International conference on web-based learning* (pp. 174–184). Springer.

Tang, B., He, H., Baggenstoss, P. M., & Kay, S. (2016). A bayesian classification approach using class-specific features for text categorization. *IEEE Transactions on Knowledge and Data Engineering, 28*(6), 1602–1606.

Tang, D., Qin, B., & Liu, T. (2015). Learning semantic representations of users and products for document level sentiment classification. In *ACL* (vol. (1), pp. 1014–1023).

Valiente, M.-C., Sicilia, M.-A., Garcia-Barriocanal, E., & Rajabi, E. (2015). Adopting the metadata approach to improve the search and analysis of educational resources for online learning. *Computers in Human Behavior, 51*, 1134–1141.

Wang, S., & Ji, Q. (2015). Video affective content analysis: A survey of state-of-the-art methods. *IEEE Transactions on Affective Computing, 6*(4), 410–430.

Wu, X., Zhu, X., Wu, G.-Q., & Ding, W. (2014). Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering, 26*(1), 97–107.

Xing, W., Chen, X., Stein, J., & Marcinkowski, M. (2016). Temporal predication of dropouts in moocs: Reaching the low hanging fruit through stacking generalization. *Computers in Human Behavior, 58*, 119–129.

Xing, C., Wang, D., Zhang, X., & Liu, C. (2014). Document classification with distributions of word vectors. In *Asia-pacific signal and information processing association, 2014 annual summit and conference (APSIPA)* (pp. 1–5). IEEE.

Yang, L., & Wang, X. (2016). Online appearance manifold learning for video classification and clustering. In *International conference on computational science and its applications* (pp. 551–561). Springer.

Please cite this article in press as: Dessì, D., et al., Bridging learning analytics and Cognitive Computing for Big Data classification in micro-learning video collections, Computers in Human Behavior (2018), https://doi.org/10.1016/j.chb.2018.03.004